

Nonparametric Empirical Bayes Growth Curve Analysis

N.S. Altman
Biometrics Unit
Cornell University
Ithaca, NY 14853

G. Casella
Biometrics Unit
Cornell University
Ithaca, NY 14853

BU-1089-MA

June 1994

Nonparametric Empirical Bayes Growth Curve Analysis

N. S. Altman and G. Casella *
Biometrics Unit
Cornell University

BU-1089-M
June 1, 1994

Abstract

Nonparametric and semiparametric regression have been suggested as alternatives to parametric models for growth curve analysis (Gasser, Müller, Köhler, Molinari, and Prader, 1984; Hart and Wehrly, 1986; Ramsay and Dalzell, 1991; Stützle et al, 1980). In this paper we demonstrate that Empirical Bayes estimation can be used to improve linear smoothing estimates when multiple curves are available.

Keywords: Shrinkage Estimation; Nonparametric Regression; Linear Smoother; Stein Estimation; Longitudinal Data; Functional Data Analysis.

*The authors were supported by NSF Grant DMS89-16245. N. Altman received support from Hatch Grant 15140 NYF. G. Casella also received support from NSF Grant No. DMS9100839 and NSA Grant 90F-073.

1 Introduction

In many situations the observed data are not scalar or multivariate measurements, but rather curves measured on each sampling unit. Examples include growth curves of organisms, learning curves, measures of curvature of the human spine, stock market trends, and studies of uptake and retention of various drugs by organisms. Analysis of such data is often called growth curve analysis. Since the data that are recorded are discretized and include measurement error, an objective of such studies is often to characterize the population of curves in some way, or to estimate the individual curves. In this paper, we focus on the use of Empirical Bayes (EB) techniques (Casella 1985, 1992; Efron and Morris 1973, 1975; Morris 1983) for improving the estimation of individual curves.

A number of techniques have been devised for analyzing growth curves. (Some reviews may be found in Geisser, 1980; Goldstein, 1979; Marubini and Milani, 1986; Rao, 1987). We find most natural those that explicitly model the curves for each sampling unit. Ramsay and Dalzell (1991) coined the term “functional data analysis” for this approach. A number of functional approaches are in common use - these include linear or polynomial regression (Geisser, 1970; Grizzle and Allen, 1969; Lee and Geisser, 1972, 1975; Potthoff and Roy, 1964; Rao, 1965), nonlinear parametric models (for example, Bock and Thissen, 1976; Gibson, Bratchell and Roberts, 1988; Jenss and Bayley, 1937), semiparametric regression models (Ramsay and Dalzell, 1991) and nonparametric regression models (Gasser et al, 1984; Hart and Wehrly, 1986; Stützel et al, 1980).

While parametric models have some attractive features (often including interpretable parameters and ease of comparison of different subgroups of the population) the parametric form is seldom derived from known mechanisms, but instead is selected arbitrarily to fit the observed data. For this reason semiparametric and nonparametric fitting, which allow adaptive fits, are competitive alternatives for growth curve modelling. The methods

used in this paper are all linear smoothers - that is, the fitted values are linear combinations of the data. Linear smoothers include a large class of commonly used techniques, including polynomial and orthogonal series regression, kernel and nearest neighbor estimators, running linear estimators and smoothing splines.

When the experimental units come from a population, it is natural to suppose that some aspects of the growth patterns will be common to the growth curves. For example, growth curves for human height show rapid growth in the first few months of life, continued, slower growth in childhood, more rapid growth during the adolescent growth spurt, and then a plateau in adulthood. The exact timing and size of these features, of course, vary from individual to individual. Parametric estimation encodes these similarities in the parametric model. Typically, nonparametric regression is applied to the individual curves, and does not utilize the similarities among curves. In this paper we utilize population information in nonparametric curve estimation by applying EB techniques to the fitted values. This is in contrast to the use of EB estimation to improve parameter estimates (rather than fitted values) for parametric growth curve models, (Berkey, 1982; Hui and Berger, 1982; Laird and Lange, 1987; Rao, 1987; Strenio, Weisberg and Bryk, 1983) .

Nonlinear and nonparametric regression estimators are, in general, biased for the true regression functions. When EB methods are used to combine unbiased estimators, the resulting estimators are biased, and the improvement in estimation is due to a bias versus variance trade-off. When EB estimators are used to combine biased estimators, the resulting estimators are recentered, which may reduce bias for some individuals. As a result, greater relative improvements in estimation error can result than in the unbiased case.

In Section 2 we discuss the growth curve model and linear smoothers. In Section 3 we discuss a shrinkage paradox for biased estimators and show that the EB formulation leads to an estimator for the shrinkage parameter. We also show heuristically that development of a Stein estimator is a

difficult problem in this context. In Section 4 we extend the EB ensemble estimator pointwise to nonparametric curve estimation and suggest a method of moments estimator of the shrinkage parameter. Section 5 summarizes a study on simulated human growth curves. Section 6 is a summary and brief discussion of the methodology.

2 Functional Approach to Growth Curve Analysis

Suppose that we have a sample of individuals from a population in which we are observing a phenomenon that changes over time (or some other dimension). This phenomenon may be some aspect of growth (if the individuals are organisms) or may be other time series such as acidification of lakes in a watershed. The important aspect is that for each individual in the population, the measurement of interest is a curve. For the i^{th} individual, the curve is

$$\mu_i(t) = \mu(t) + \eta_i(t)$$

where t is time, $\mu(t)$ is the mean curve over the whole population, and $\eta_i(t)$ is the deviation of the individual's curve from the mean curve at time t .

The population is sampled to obtain a set of N individuals, and each individual is observed at times t_1, t_2, \dots, t_T . The methods discussed in this article require that individuals have been measured at a fixed set of design points; however simulation results in Section 5 demonstrate that the methods can accommodate missing values. The data on the i^{th} individual are $y_{i,1}, \dots, y_{i,T}$ where

$$y_{i,j} = \mu_i(t_j) + \varepsilon_{i,j}. \tag{1}$$

Nonparametric regression methods for growth curves assume that $\mu_i(t)$ is smooth (that is, $\mu_i(t)$ is continuous and differentiable), and the errors $\varepsilon_{i,j}$ have mean zero and constant variance σ^2 . The goal is to recover $\mu_i(t)$ and

a number of techniques are available. In this article, we limit ourselves to linear smoothers. That is, the estimate, $\hat{\mu}_i(t)$ has the form:

$$\hat{\mu}_i(t) = \sum_{j=1}^T w_\lambda(t, j) y_{i,j} \quad (2)$$

where the weight function, $w_\lambda(t, j)$ depends on the observation times, $t_1 \cdots t_T$, the estimation time, t , and a smoothing parameter, λ , which controls the smoothness of the estimator. Some examples of this type of estimator include polynomial regression (λ =degree), truncated Fourier series (λ =highest frequency), kernel regression (λ =bandwidth), smoothing splines (λ =smoothing parameter) and running linear regression (λ =span). A good review that provides some unifying ideas on these types of smoothers is Eubank (1988).

Linear smoothers have in common the following features which are important to the use of EB estimation:

1. The estimators are biased (except for special cases). That is:

$$\begin{aligned} E[\hat{\mu}_i(t) | \mu_i(\bullet)] &= \sum_{j=1}^T w_\lambda(t, j) \mu_i(t_j) \\ &\neq \mu_i(t). \end{aligned}$$

($\mu_i(\bullet)$ denotes the entire function.)

2. $Var[\hat{\mu}_i(t_j) | \mu_i(\bullet)] < Var[y_{i,j} | \mu_i(\bullet)]$. That is, smoothing reduces the variance compared to the unbiased estimator $y_{i,j}$.
3. If the data are Normally distributed, then $(y_{1,1} \cdots y_{N,T}, \hat{\mu}_1(t_1) \cdots \hat{\mu}_N(t_T))$ is jointly Normal.

3 Empirical Bayes Estimation

In this section we attempt to learn a little more about the form and properties of an EB estimator appropriate for nonparametric regression estimation. Since linear smoothers are biased, we first address the simpler problem of

developing an EB estimator for a univariate problem when the individual estimators are biased. This estimator will be extended to a curve estimator in Section 4. We also look at some of the risk properties of our estimator in a simplified setting and find that, although risk domination of the individual estimators should be expected, development of a dominating estimator is difficult.

3.1 Constructing an EB Estimator from Biased Estimators

To understand how EB estimation can be used for curve estimation, it is useful to first look at a univariate case with a biased estimator. Suppose that we observe $z_{i,j}$ where for each fixed i , $z_{i,j}$ are independent and identically distributed with cumulative distribution function $F(\bullet|\mu_i)$. Suppose that μ_i is estimated by

$$\hat{\mu}_i = \hat{\mu}_i(z_{i,1} \cdots z_{i,T})$$

using data only from individual i . We evaluate the “goodness” of the estimates using squared error loss (SE), where the SE of the estimator $(\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_N)$, as an estimate of $(\mu_1, \mu_2, \dots, \mu_N)$, is given by

$$SE = \sum_{i=1}^N (\hat{\mu}_i - \mu_i)^2.$$

Typically, the estimators $\hat{\mu}_i$ are biased and can be written in the form:

$$\hat{\mu}_i = \psi_i + \nu_i$$

where ψ_i is an unknown parameter with $E(\hat{\mu}_i) = \psi_i$, and ν_i is an error term with $E(\nu_i) = 0$ and $Var(\nu_i) = \tau^2/T$.

Let $\bar{\mu} = \frac{1}{N} \sum_{i=1}^N \mu_i$ and $\bar{\hat{\mu}} = \frac{1}{N} \sum_{i=1}^N \hat{\mu}_i$. Then there is an α such that

$$\tilde{\mu}_i = \bar{\mu} + \alpha(\hat{\mu}_i - \bar{\hat{\mu}}) \tag{3}$$

has lower SE than $\hat{\mu}_i$. Simple algebra shows that the optimal value is

$$\alpha = S_{\hat{\mu}\mu} / S_{\hat{\mu}\hat{\mu}} \tag{4}$$

where we define $S_{XY} = \sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})/(N-1)$. This is an example of a Stein paradox for biased estimators in that $\tilde{\mu}_i$ uses information from the other populations, which may not be related in any way to the i^{th} population. (See Green and Strawderman, 1991, for another example of Stein estimation with biased estimators.)

When $\psi_i = \mu_i$, $\hat{\mu}_i$ is unbiased for μ_i , and $0 \leq \alpha \leq 1$. The result is well-known for this case (James and Stein, 1961; Stein, 1956) and α is called the shrinkage parameter. We will continue to use this term even though, when $\hat{\mu}_i$ is biased for μ_i , α can be negative or greater than 1.

Since the denominator is positive, α is negative only if $S_{\hat{\mu}\mu} < 0$. The term $\alpha(\hat{\mu}_i - \bar{\mu})$ transforms the smallest $\hat{\mu}_i$ into the largest $\tilde{\mu}_i$ and so on, reversing the ordering of the estimates so that they are positively correlated with the estimands. Similarly α can be greater than 1 only if $S_{\mu\mu} > S_{\hat{\mu}\mu}$, that is, only if the μ_i 's are more spread out than the $\hat{\mu}_i$'s. The term $\alpha(\hat{\mu}_i - \bar{\mu})$ then increases the spread of the estimates. In both cases, α moves the ensemble estimators, $\tilde{\mu}_i$, to behave more like the collection of μ_i 's.

As with the Stein paradox for unbiased estimators, the result can be re-expressed in Bayesian terms using the hierarchical model:

$$\hat{\mu}_i | \psi_i \sim \text{Normal}(\psi_i, \tau^2/T)$$

$$\begin{pmatrix} \psi_i \\ \mu_i \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \psi \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_\psi^2 & \sigma_{\psi\mu} \\ \sigma_{\psi\mu} & \sigma_\mu^2 \end{pmatrix}\right)$$

where, for now, we assume that the variances and covariances are known. Since μ_i is the parameter to be estimated, we can rewrite the model by integrating out ψ_i . This yields

$$\begin{pmatrix} \hat{\mu}_i \\ \mu_i \end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix} \psi \\ \mu \end{pmatrix}, \begin{pmatrix} \sigma_\mu^2 & \sigma_{\psi\mu} \\ \sigma_{\psi\mu} & \sigma_\mu^2 \end{pmatrix}\right) \quad (5)$$

where $\sigma_\mu^2 = \sigma_\psi^2 + \tau^2/T$. The posterior mean of μ_i , a Bayes estimator, can be written

$$\begin{aligned} \tilde{\mu}_i &= E(\mu_i | \hat{\mu}_i) \\ &= \mu + B(\hat{\mu}_i - \psi) \end{aligned} \quad (6)$$

where

$$B = \sigma_{\psi\mu}/\sigma_{\tilde{\mu}}^2. \quad (7)$$

Note that (6) and (7) are population versions of (3) and (4). The estimator $\tilde{\mu}_i$ in (3) is a candidate on which to base an EB estimator arising from the hierarchical argument.

The relative mean squared error (RMSE) of $\tilde{\mu}_i$ versus $\hat{\mu}_i$ is

$$\frac{\sigma_{\mu}^2 - B\sigma_{\psi\mu}}{\sigma_{\psi\mu}/B + \sigma_{\mu}^2 - 2\sigma_{\psi\mu} + (\mu - \psi)^2}$$

For the unbiased case ($\mu_i = \psi_i$ and $\sigma_{\mu}^2 = \sigma_{\psi\mu}$), the RMSE is just B . Table 1 shows the potential gains in RMSE when $\mu = \psi$ (which is clearly the case least favorable to the use of $\tilde{\mu}_i$). (Without loss of generality we assume $\sigma_{\mu}^2 = 1$.) Shrinkage is most effective when the $\hat{\mu}_i$'s are poorly determined or their conditional expectations (ψ_i 's) are more spread out than the μ_i 's. Even when the mean and variance of ψ_i 's and the μ_i 's are the same, greater gains in mean squared error are possible in the biased than in the unbiased case, suggesting that an EB approach may lead to large savings in SE.

[PLACE TABLE 1 ABOUT HERE]

3.2 A Heuristic Approach to Shrinkage

In the unbiased case, it is well-known that use of the method of moments to estimate the unknown parameters in the hierarchy usually improves the SE, but is not guaranteed to do so. Instead, a number of estimators have been proposed which are generally called “shrinkage” or Stein estimators. (For a recent review, see Brandwein and Strawderman, 1990.)

For the biased case we have not been able to devise a shrinkage estimator. However, the following heuristics provide some insight into the problem.

In model (5), take $\mu = \psi = 0$. (Since these are values specified a priori, this is not a loss of generality.) Then

$$\hat{\mu}_i|\mu_i \sim Normal(a\mu_i, \tau^{*2}),$$

where $a = \sigma_{\psi\mu}/\sigma_\mu^2$ and $\tau^{*2} = \sigma_{\psi\mu}/B - \sigma_{\psi\mu}^2/\sigma_\mu^2$. If we assume τ^{*2} is known and equal to 1, we reduce consideration to the simpler stylized problem:

Observe independent random variables X_1, \dots, X_N , where

$X_i \sim \text{Normal}(a\mu_i, 1)$.

Find a good estimator of $\mu = (\mu_1, \dots, \mu_N)'$.

If we assume a is known, then the usual estimator of μ is \mathbf{X}/a , and the Stein estimator

$$\delta_a(\mathbf{X}) = \left(1 - \frac{c}{|\mathbf{X}|^2}\right) \frac{\mathbf{X}}{a},$$

dominates \mathbf{X}/a if $0 < c < 2(N-2)$ where $|\mathbf{X}|^2 = \Sigma \mathbf{X}_i^2$. In almost all cases of interest, however, we must consider a to be unknown and estimated by a quantity \hat{a} . The natural Stein estimator to consider is then analogously

$$\delta_{\hat{a}}(\mathbf{X}) = \left(1 - \frac{c}{|\mathbf{X}|^2}\right) \frac{\mathbf{X}}{\hat{a}}, \quad (8)$$

Our first goal is to compare the mean squared error ($\text{MSE} = E(\text{SE})$) of $\delta_{\hat{a}}(\mathbf{X})$ and \mathbf{X}/\hat{a} , the “usual” estimator of μ . Calculating the difference in MSE yields (details are in the Appendix):

$$\begin{aligned} & \text{MSE}(\delta_{\hat{a}}) - \text{MSE}(\mathbf{X}/\hat{a}) \\ &= E \left\{ \left[\frac{1}{\hat{a}^2} \right] \left[\frac{c}{|\mathbf{X}|^2} [c - 2\frac{\hat{a}}{a}(N-2)] - 2c \left(1 - \frac{\hat{a}}{a}\right) \right] \right\} \\ & \quad + \frac{2c}{a} E \left(\frac{1}{\hat{a}^2 |\mathbf{X}|^2} \sum_{i=1}^N X_i \frac{\partial}{\partial X_i} \hat{a} \right). \end{aligned} \quad (9)$$

Although (9) looks complicated, it does yield a lot of information about possible dominance. Firstly, if the estimator \hat{a} is independent of \mathbf{X} , the second expectation is zero, due to the derivative. (This situation is probably most likely, since we have only N observations to use in estimating the N means plus a . Thus an estimator of a must be based on other information which may be independent of the observations.) If \hat{a} is not independent of \mathbf{X} , then we would want the second expectation in (9) to be negative. If \hat{a} is

a function of $|\mathbf{X}|^2$, say $\hat{a} = \hat{a}(|\mathbf{X}|^2)$, then $\sum_{i=1}^N X_i \frac{\partial}{\partial X_i} \hat{a} = 2|\mathbf{X}|^2 \frac{d}{d|\mathbf{X}|^2} \hat{a}(|\mathbf{X}|^2)$ so \hat{a} should be decreasing in $|\mathbf{X}|^2$.

If \hat{a} is a consistent estimator of a , then the condition for dominance of the Stein estimator for known a will take over, and $\delta_{\hat{a}}$ will dominate \mathbf{X}/\hat{a} if $c < 2(N - 2)$. Thus it is reasonable to assume that $\delta_{\hat{a}}$ will perform better than \mathbf{X}/\hat{a} as long as \hat{a} is a reasonable estimator of a .

As a last question, one can inquire whether \mathbf{X}/\hat{a} is, in fact, a better estimator than \mathbf{X} . If not, domination of the latter should be the goal. Calculations give somewhat inconclusive results. Superiority of \mathbf{X}/\hat{a} over \mathbf{X} depends on $|\mu|^2$. However, if \hat{a} is a consistent estimator, then \mathbf{X}/\hat{a} is preferred to \mathbf{X} .

In the growth curve problem, estimates of a are available due to the longitudinal component of each curve. However, as T increases, adaptive choice of smoothing parameter allows both the bias and variance of the curve estimators to diminish, reducing the need for shrinking. Clearly developing a Stein-type estimator for this problem, one that is guaranteed to dominate the individual estimators in total risk, is a difficult task. However, we have seen that good risk improvement is possible. Since we expect our EB estimators to behave similarly to their Stein-type counterparts, we expect them to yield good risk improvement.

4 Linear Smoothing and Empirical Bayes Estimation

We now return to the EB approach to the problem of ensemble nonparametric curve estimation.

In the smoothing problem, the data are described by equation (1) and the estimator of the mean function by equation (2). Note that $\hat{\mu}_i(t)$ is biased for $\mu_i(t)$ and Stein's paradox holds for the point estimator $\hat{\mu}_i(t)$. When the data are Normally distributed, so is $\hat{\mu}_i(t)$ for fixed smoothing parameter, so that under the assumption that $\psi_i(t) = E[\hat{\mu}_i(t)|\mu_i(\bullet)]$ and $\mu_i(t)$ are jointly

Normally distributed we have, as in model (5)

$$\begin{pmatrix} \hat{\mu}_i(t) \\ \mu_i(t) \end{pmatrix} \sim Normal \left(\begin{pmatrix} \psi(t) \\ \mu(t) \end{pmatrix}, \begin{pmatrix} \sigma_{\hat{\mu}(t)}^2 & \sigma_{\psi(t)\mu(t)} \\ \sigma_{\psi(t)\mu(t)} & \sigma_{\mu(t)}^2 \end{pmatrix} \right) \quad (10)$$

where $E[\psi_i(t)] = \psi(t)$, $\sigma_{\hat{\mu}(t)}^2$ is the unconditional variance of the estimators, $\sigma_{\psi(t)\mu(t)}$ is the covariance between the $\psi_i(t)$'s and the $\mu_i(t)$'s, and $\sigma_{\mu(t)}^2$ is the variance of the $\mu_i(t)$'s.

Note that $\sigma_{\hat{\mu}(t)}^2 = \sigma_{\psi}^2(t) + Var[\hat{\mu}_i(t)|\mu_i(\bullet)]$ and that typically for non-parametric smoothers, $TVar[\hat{\mu}_i(t)|\mu_i(\bullet)]$ is asymptotic to a function of λ , so that $\sigma_{\hat{\mu}(t)}^2 \approx \sigma_{\psi}^2(t) + \tau^2/T$ where $TVar[\hat{\mu}_i(t)|\mu_i(\bullet)] \rightarrow \tau^2$. Thus for fixed t the model is very similar to model (5) although, for each i , the data were generated by a regression model rather than random sampling.

Following the development in Section 3.1, we obtain (using Equation 6)

$$\begin{aligned} \tilde{\mu}_i(t) &= E[\mu_i(t)|\hat{\mu}_i(\bullet)] \\ &= \mu(t) + B(t)(\hat{\mu}_i(t) - \psi(t)) \end{aligned} \quad (11)$$

where $B(t) = \sigma_{\psi(t)\mu(t)}/\sigma_{\hat{\mu}(t)}^2$. Allowing the Bayesian shrinkage parameter to depend on t allows for the fact that both the bias and variance of a linear smoother vary with t .

All of the unknown parameters in (11) can be estimated at the design points using the method of moments. Since $\mu(t_j) = E(y_{i,j})$ and $\psi(t_j) = E[\hat{\mu}_i(t_j)]$ natural estimators of these parameters are the respective sample means $\bar{y}_{\bullet,j}$ and $\bar{\hat{\mu}}(t_j)$ defined in Table 2. A natural estimator of $\sigma_{\hat{\mu}(t_j)}^2$ is the sample variance, $\hat{\sigma}_{\hat{\mu}(t_j)}^2$ also defined in Table 2. Finally, $Cov[y_{i,j}, \hat{\mu}_i(t_j)] = \sigma_{\psi(t_j)\mu(t_j)} + w_{\lambda}(t_j, j)\sigma^2$ so that a natural estimator of $\sigma_{\psi(t_j)\mu(t_j)}$ is of the form $\widehat{Cov}[y_{i,j}, \hat{\mu}_i(t_j)] - w_{\lambda}(t_j, j)\hat{\sigma}^2$, where $\widehat{Cov}[y_{i,j}, \hat{\mu}_i(t_j)]$ is the sample covariance of the y 's and the estimators, and $\hat{\sigma}^2$ is an estimator of the residual variance. A number of estimators of the residual variance are available (Buckley, Eagleson and Silverman, 1988; Gasser, Sroka, Jennen-Steinmetz, 1986; Rice, 1984). Once an estimator $\hat{\sigma}^2$ has been selected, $\hat{\sigma}_{\psi(t_j)\mu(t_j)}$ can be defined as in Table 2.

[PLACE TABLE 2 ABOUT HERE]

An initial estimate of $B(t_j)$ is computed using the method of moments:

$$\hat{B}(t_j) = \frac{\hat{\sigma}_{\psi(t_j)\mu(t_j)}}{\hat{\sigma}_{\hat{\mu}}^2(t_j)}.$$

Since $\hat{B}(t_j)$ is very noisy, we suggest the estimates be smoothed to new estimates $b(t_j)$. The final estimator at the design points is then

$$\tilde{\mu}_i(t_j) = \bar{y}_{\bullet,j} + b(t_j)[\hat{\mu}_i(t_j) - \bar{\hat{\mu}}(t_j)]. \quad (12)$$

The only term of this equation that is not defined at general points t is $\bar{y}_{\bullet,j}$. However, by interpolating this term between design points, the Empirical Bayes estimator can be extended to general points. These computations have been summarized in Table 2.

5 Simulation Results

To determine the efficacy of the empirical Bayes estimator in a realistic setting, 3 small simulation studies were performed. The first study explores the effects of varying the number of curves and the number of design points per curve. The second study explores the effects of varying the within curve measurement error. The third study explores the effects of missing data.

Growth curves were generated from the Bock and Thissen model (Bock and Thissen, 1976) for human growth (height):

$$\mu_i(t) = \sum_{k=1}^3 \frac{A_{i,k}}{1 + \exp[-C_{i,k}(t - D_{i,k})]}.$$

The values of $A_{i,k}$, $C_{i,k}$, $D_{i,k}$ were generated as multivariate normals with means and covariance functions computed from the Fels growth data for boys (Bock, personal communication, 1992). For each i , a mean curve was generated for ages 5 to 20. Only mean curves resulting in height at age 20 between 157.5 and 193 cm. were used (covering over 95% of the American adult population in 1976-1980 as reported in Freedman, Pisani, Purves and

Adhikari, 1991), resulting in use of about 90% of the generated curves in most cases. Normal errors were added to the mean curves to obtain the “data”.

Smoothing was done using the cubic smoothing spline estimator “bart” available from netlib (O’Sullivan, 1985). The smoothing parameter was chosen using generalized cross-validation (Craven and Wahba, 1979) averaged over all the curves in the set.

The method of moments estimator of the shrinkage parameter, described in Section 4 was computed at each design point, using the Gasser – Sroka – Jennen-Steinmetz estimator of residual variance averaged over the curves. The shrinkage parameters were smoothed (also using the cubic smoothing spline with smoothing parameter chosen by generalized cross-validation) and the final estimates were computed from equation (12).

In the first study, sets of 20 or 50 different mean curves were generated according to the Bock and Thissen model. Normal errors with standard deviation 4 centimeters were added to the mean curves satisfying the inclusion criterion to obtain the “observed” growth data. Each simulation consisted of 100 replications.

Figure 1 displays 5 results from a set of 50 curves (of which 43 were in the target range) with 8 design points per curve, and the smoothed curve of shrinkage values for the set. Plots 1a-1e show the difference between the true mean curve and the spline or EB estimate. Note that there is no evidence that the EB estimates are being shifted. Figure 1f shows that this is due to the narrow range of the shrinkage values which are very close to 1.0. However, it is evident in all the plots that the Empirical Bayes estimator is being pulled away from the spline estimator in the mid-teen years.

For the replicate displayed in Figure 1, Figure 2 displays the errors averaged over the 43 curves. The figure shows clearly that the kernel estimator is biased for height during mid-teen years. The errors from the Empirical Bayes estimator are much smaller in this region. The Empirical Bayes estimator is achieving improvements in SE by recentering the estimates in the

mid-teen years, rather than by “shrinking” the curves.

Figure 3 displays the ratio of SE of the spline and EB estimates for the 100 replications of each of the 4 simulations. The effectiveness of the Empirical Bayes estimator is greater when there are more curves per set, and when there are fewer design points per curve. With 16 points on each curve, and a sample size of only 20, the EB estimates improved the SE in 54 of 100 trials. With a sample size of 50, the EB estimates improved the SE in 90 of 100 trials. With only 8 points on each curve and a sample size of 50, the EB estimates improved the SE in all but 1 trial.

The measurement error of 4 cm. is unrealistically large for human growth curves. Buckler (1979) reports deviations as large as 2.8 cm. over the course of a day in adolescent boys participating in growth studies, but deviations of less than 1 cm. are more typical. However the ratio of measurement error to deviation among curves is realistic for many examples. The second simulation study explores the effects of measurement error on the EB adjustment. Sets of 50 curves with 8 points per curve were generated. Normal errors with standard deviations of 1, 2 or 4 cm. were added to the mean curves. The resulting changes in relative SE are summarized in Figure 4. When the standard deviation of the normal errors is 4 cm., the EB estimates improved the SE in all but 1 trial. When the standard deviation is reduced to 2 cm., the EB estimates improved the SE in 74 of 100 trials, but when the standard deviation is reduced to 1 cm., there was improvement in only 1 trial. Sample sizes of several hundred are required to attain improvements at this level of measurement error. This is not surprising. The standard deviation of “true” heights at age 20 is about 7.6 cm. Comparing with Table 1, this is approximately the case $\tau^2 = .02$.

The final study explored the effects of missing values on the EB estimator. A Bernoulli process was generated to remove 25% of the data at random, from sets of 50 curves with 8 points per curve and error standard deviation 4 cm. Curves were not used if more than 2 points were missing. To adjust for this, 60 rather than 50 curves were generated per set. The

sample sizes for missing and non-missing data were therefore not exactly the same - the study with no missing data averaged 45 curves fitting the inclusion criteria, while those with missing data averaged 49 - 50 curves.

The method of moments estimators were adjusted simply by using all available data. The estimator of error variance was adjusted by using only adjacent sets of 3 non-missing points. This leads to a much smaller sample size for error variance than for the other statistics required.

The results are displayed in Figure 5. Missing values did lead to some degradation of the EB estimator, despite the somewhat larger samples. Missing values would lead to less accurate estimation of the smoothing parameter for the spline estimates, of the spline fits, and of the shrinkage parameters. For this study, in which the fraction missing was quite large, these factors appear to have somewhat balanced out in terms of the relative improvement of SE achieved by the EB estimator. (Of course the SE increases as the amount of data decreases.)

6 Conclusions

Empirical Bayes estimation provides a means of combining evidence across similar populations. The classical development of the estimator is in the context of unbiased parametric estimators. In this context, EB methodology has been used to improve the estimation of growth parameters in polynomial growth models.

In this article we have shown that EB methodology can be extended to biased estimators, and may be even more effective in this context. We have applied these EB techniques pointwise to the output of a linear smoother, using naive method of moments estimators to estimate the shrinkage parameters. Improvement is greater when the within curve error is larger, the number of design points per curve is smaller or the sample of curves is larger. Thus the methodology will be most useful for studies in which the data are noisy (such as microbial growth) and less useful for studies in

which the within curve measurement error is small compared to the between curve variation (such as human height) or in which the curve can be sampled intensively. The method can readily be extended to studies in which all experimental units were measured at the same design points, but for which some values are missing.

7 Appendix

Derivation of Equation (9) relies on application of Stein's Identity (Stein, 1981 or Stigler, 1990). If $\mathbf{Y} = (Y_1, \dots, Y_N)$ is distributed $N_N(\eta, I)$ where I is the $N \times N$ identity matrix then

$$E[g(\mathbf{Y})\eta] = E[g(\mathbf{Y})\mathbf{Y}] - \sum_{i=1}^N E\left[\frac{\partial}{\partial Y_i} g(\mathbf{Y})\right]$$

provided all the expectations exist. The MSE of $\delta_{\hat{a}}$ is given by

$$\begin{aligned} MSE(\delta_{\hat{a}}) &= E|\delta_{\hat{a}} - \mu|^2 \\ &= E\left|\left(1 - \frac{c}{|\mathbf{X}|^2}\right) \frac{\mathbf{X}}{\hat{a}} - \mu\right|^2 \\ &= E\left|\left(\frac{\mathbf{X}}{\hat{a}} - \mu\right) - \frac{c\mathbf{X}}{\hat{a}|\mathbf{X}|^2}\right|^2 \\ &= MSE\left(\frac{\mathbf{X}}{\hat{a}}\right) - 2cE\left[\left(\frac{\mathbf{X}}{\hat{a}} - \mu\right)' \frac{\mathbf{X}}{\hat{a}|\mathbf{X}|^2}\right] + c^2E\left(\frac{1}{\hat{a}^2|\mathbf{X}|^2}\right) \end{aligned} \quad (13)$$

We now apply Stein's Identity to the middle term in (13) with $g(\mathbf{X}) = \mathbf{X}/(\hat{a}|\mathbf{X}|^2)$. Recalling that the mean of \mathbf{X} is $a\mu$, we have

$$E\left(\frac{\mathbf{X}'\mu}{\hat{a}|\mathbf{X}|^2}\right) = \frac{1}{a}E\left[\frac{1}{\hat{a}}\left(1 - \frac{N-2}{|\mathbf{X}|^2}\right)\right] + \frac{1}{a}E\left[\frac{1}{\hat{a}^2|\mathbf{X}|^2} \sum_{i=1}^N X_i \frac{\partial}{\partial X_i} \hat{a}\right] \quad (14)$$

Substituting (14) into (13) and rearranging then gives (9).

8 References

Berkey, C.S. (1982) "Bayesian Approach for a Nonlinear Growth Model," *Biometrics*, **38**, 953-961.

- Bock, R.D. and Thissen, D. (1976) "Fitting Multi-component Models for Growth in Stature," *Proceedings of the 9th International Biometrics Conf.*, Biometrics Society, Boston.
- Brandwein, A.C. and Strawderman, W.E. (1990) "Stein Estimation: The Spherically Symmetric Case," *Statistical Science*, **5**, 356-369.
- Buckler, J. M. H. (1979) "Variations in Height Throughout the Day," *Archives of Disease in Childhood*, **53**, 762.
- Buckley, M.J., Eagleson, G.K., Silverman, B. W., 1988 "The Estimation of Residual Variance in Nonparametric Regression," *Biometrika*, **75**, 189-199.
- Casella, G. (1985) "An Introduction to Empirical Bayes Data Analysis," *American Statistician*, **39**, 83-87.
- (1992) "Illustrating Empirical Bayes Methods," *Chemometrics and Intelligent Laboratory Systems* **16**, 107-125.
- Craven, P. and Wahba, G. (1979) "Smoothing Noisy Data With Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numerische Mathematika*, **31**, 377-403.
- Efron, B. and Morris, C. (1973) "Stein's Estimation Rule and Its Competitors - An Empirical Bayes Approach," *Journal of the American Statistical Association*, **68**, 117-130.
- (1975) "Data Analysis Using Stein's Estimator and Its Generalizations," *Journal of the American Statistical Association*, **70**, 311-331.
- Eubank, R. L. (1988) *Spline Smoothing and Nonparametric Regression*, New York: Marcel Dekker.
- Freedman, D., Pisani, R., Purves, R., and Adhikari, A. (1991) *Statistics*, 2nd Ed., New York: W. W. Norton and Co.
- Gasser, T., Müller, H-G., Köhler, W., Molinari, L. and Prader, A. (1984) "Nonparametric Regression Analysis of Growth Curves," *Annals of Statistics*, **12**, 210-229.
- Gasser, T., Sroka, L., and Jennen-Steinmetz, C. (1986) "Residual Variance and Residual Pattern in Nonlinear Regression," *Biometrika*, **73**, 625-633.

- Geisser, S. (1970) "Bayesian Analysis of Growth Curves," *Sankhyā Ser. A*, **32**, 53-64.
- (1980) "Growth Curve Analysis," in *Handbook of Statistics, Vol. 1*, Krishnaiah, P.R. (ed.), Amsterdam: North-Holland, pp. 89-115.
- Gibson, A.M., Bratchell, N., and Roberts, T.A. (1988) "Predicting Microbial Growth: Growth Responses of Salmonellae in a Laboratory Medium as Affected by pH, Sodium Chloride and Storage Temperature," *International Journal of Food Microbiology*, **6**, 155-178.
- Goldstein, H. (1979) *The Design and Analysis of Longitudinal Studies*, London: Academic Press.
- Green, E. J. and Strawderman, W. E. (1991) "A James-Stein Type Estimator for Combining Unbiased and Possibly Biased Estimators," *Journal of the American Statistical Association*, **86**, 1001-1006.
- Grizzle, J. E. and Allen, D. M. (1969) "Analysis of Growth and Dose Response Curves," *Biometrics*, **25**, 357-382.
- Hart, J.D. and Wehrly, T.E. (1986) "Kernel Regression Estimation Using Repeated Measurements Data," *Journal of the American Statistical Association*, **81**, 1080-1088.
- Hui, S.L. and Berger, J.O. (1983) "Empirical Bayes Estimation of Rates in Longitudinal Studies," *Journal of the American Statistical Association*, **78**, 753-760.
- James, W. and Stein, C. (1961) "Estimation With Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium Mathematical Statistics and Probability*, **1**, 361-380, Berkeley: University of California Press.
- Jenss, R.M. and Bayley, N. (1937) "A Mathematical Method for Studying the Growth of a Child," *Human Biology*, **9**, 556-563.
- Laird, N. and Lange, N. (1987) Comment on "Predictions of Future Observations in Growth Curve Models" by C. R. Rao, *Statistical Science*, **2**, 451-454.
- Lee, J. C. and Geisser, S. (1972) "Growth Curve Prediction," *Sankhyā Ser. A*, **34**, 394-412.

- (1975) “Applications of Growth Curve Prediction,” *Sankhyā Ser. A*, **37**, 239-256.
- Marubini, E. and Milani, S. (1986) “Approaches to the Analysis of Longitudinal Data,” in *Human Growth: A Comprehensive Treatise* Vol. 3, Falkner, F., Tanner, J.M. (eds.) New York: Plenum Press, pp. 79-94.
- Morris, C. (1983) “Parametric Empirical Bayes Inference: Theory and Applications,” *Journal of the American Statistical Association*, **78**, 47-65.
- O’Sullivan, F. (1985) Comment on “Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting,” by B. W. Silverman, *Journal of the Royal Statistical Society Series B*, **47**, 39-40.
- Potthoff, R.F. and Roy, S.N. (1964) “A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems,” *Biometrika*, **51**, 313-326.
- Ramsay, J. O. and Dalzell, C. J. (1991) “Some Tools for Functional Data Analysis,” (with comments) *Journal of the Royal Statistical Society. Ser. B*, **53**, 539-572.
- Rao, C.R. (1965) “The Theory of Least Squares When the Parameters are Stochastic and Its Application to the Analysis of Growth Curves,” *Biometrika*, **52**, 447-458.
- (1987) “Predictions of Future Observations in Growth Curve Models,” (with comments) *Statistical Science*, **2**, 434-471.
- Rice, J. (1984) “Bandwidth Choice for Nonparametric Regression,” *Annals of Statistics*, **12**, 1215-1230.
- Stein, C. (1956) “Inadmissibility of the Usual Estimator of the Mean of a Multivariate Normal Distribution,” *Proceedings of the Third Berkeley Symposium Mathematical Statistics and Probability*, **1**, 197-206. Berkeley: University of California Press.
- (1981) “Estimation of the Mean of a Multivariate Normal Distribution,” *Annals of Statistics*, **9**, 1135-1151.
- Stigler, S. (1990) “A Galtonian Perspective on Shrinkage Estimation,” *Statistical Science*, **5**, 147-155.

- Strenio, J.F., Weisberg, H.I., Bryk, A.S. (1983) "Empirical Bayes Estimation of Individual Growth-Curve Parameters and Their Relationship to Covariates," *Biometrics*, **39**, 71-86.
- Stützle, W., Gasser, Th., Molinari, L., Largo, R.H., Prader, A., and Huber, P.J. (1980) "Shape-invariant Modeling of Human Growth," *Annals of Human Biology*, **7**, 507-528.

		unbiased case	biased case $\psi = \theta$		
			$\sigma_{\psi\theta} = .8\sigma_\psi$		
τ^2	T	$RMSE = B$	σ_ψ^2	B	$RMSE$
.25	6	.96	.25	1.37	.92
			1.0	.77	.87
			4.0	.40	.20
	30	.99	.25	1.55	.83
			1.0	.79	.89
			4.0	.40	.20
1.0	6	.86	.25	.96	1.00
			1.0	.69	.80
			4.0	.38	.20
	30	.97	.25	1.41	.90
			1.0	.77	.88
			4.0	.40	.20
4.0	6	.60	.25	.44	.74
			1.0	.48	.58
			4.0	.34	.18
	30	.88	.25	1.04	1.00
			1.0	.71	.82
			4.0	.20	.39

Table 1: Relative mean squared error of $\tilde{\mu}_i$ versus $\hat{\mu}_i$ when $\psi = \mu$, $\sigma_\mu^2 = 1$ and the Bayes parameters are known. Shrinkage is most effective when the $\hat{\mu}_i$'s are poorly determined or their conditional expectations (ψ_i 's) are more spread out than the μ_i 's. Even when the mean and variance of ψ_i 's and the μ_i 's is the same, greater gains in MSE are possible in the biased than in the unbiased case.

Section 3.1 Quantity	Section 4 Quantity	Estimated by
μ	$\mu(t_j)$	$\bar{y}_{\bullet,j} = \frac{1}{N} \sum_{i=1}^N y_{i,j}$
ψ	$\psi(t_j)$	$\bar{\mu}(t_j) = \frac{1}{N} \sum_{i=1}^N \hat{\mu}_i(t_j)$
$\sigma_{\hat{\mu}}^2$	$\sigma_{\hat{\mu}(t_j)}^2$	$\hat{\sigma}_{\psi(t_j)}^2 = \frac{1}{N-1} \sum_{i=1}^N [\hat{\mu}_i(t_j) - \bar{\mu}(t_j)]^2$
$\sigma_{\psi\mu}$	$\sigma_{\psi(t_j)\mu(t_j)}$	$\hat{\sigma}_{\psi(t_j)\mu(t_j)} = \frac{1}{N} \sum_{i=1}^N (y_{i,j} - \bar{y}_{\bullet,j})[\hat{\mu}_i(t_j) - \bar{\mu}(t_j)] - w_{\lambda}(t_j, j)\hat{\sigma}^2$
B	$B(t_j)$	$\hat{B}(t_j) = \hat{\sigma}_{\psi(t_j)\mu(t_j)} / \hat{\sigma}_{\hat{\mu}(t_j)}^2$
B	$B(t)$	$b(t)$ obtained by smoothing $\hat{B}(t_j)$
$\tilde{\mu}_i$	$\tilde{\mu}_i(t_j)$	$\tilde{\mu}_i(t_j) = \bar{y}_{\bullet,j} + b(t)[\hat{\mu}_i(t_j) - \bar{\mu}(t_j)]$

Table 2: Method of Moments Estimators of Empirical Bayes Parameters

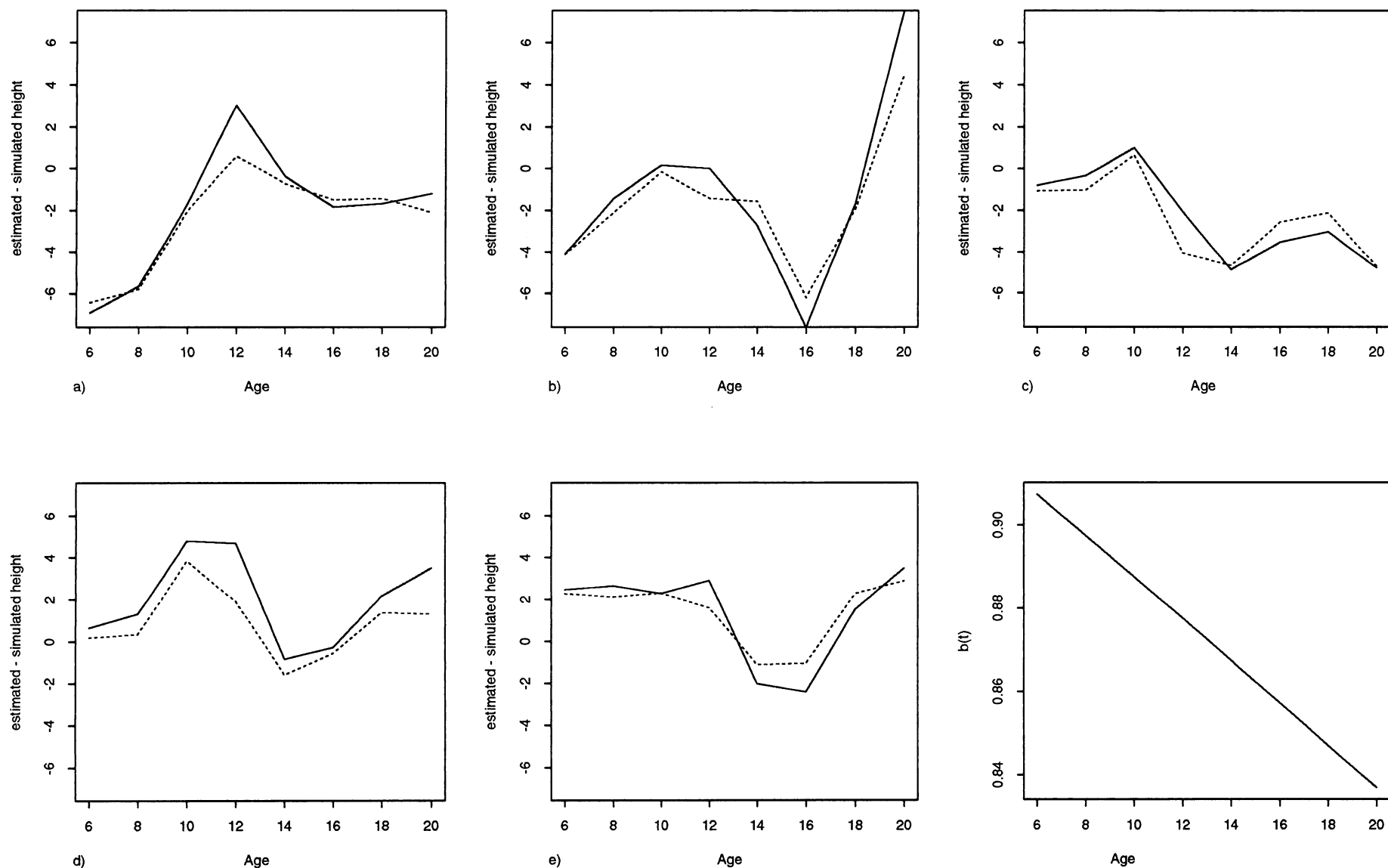


Figure 1: Plots a-e are the differences between the simulated growth curves and the smoothing spline estimates (solid) and EB estimates (dashed) for 5 of a set of 43 curves with 8 design points per curve. Note that the EB differences are not consistently above or below the spline estimates, indicating that little shrinkage is taking place. Plot f) displays the estimated shrinkage values, $b(t)$, which are close to 1.

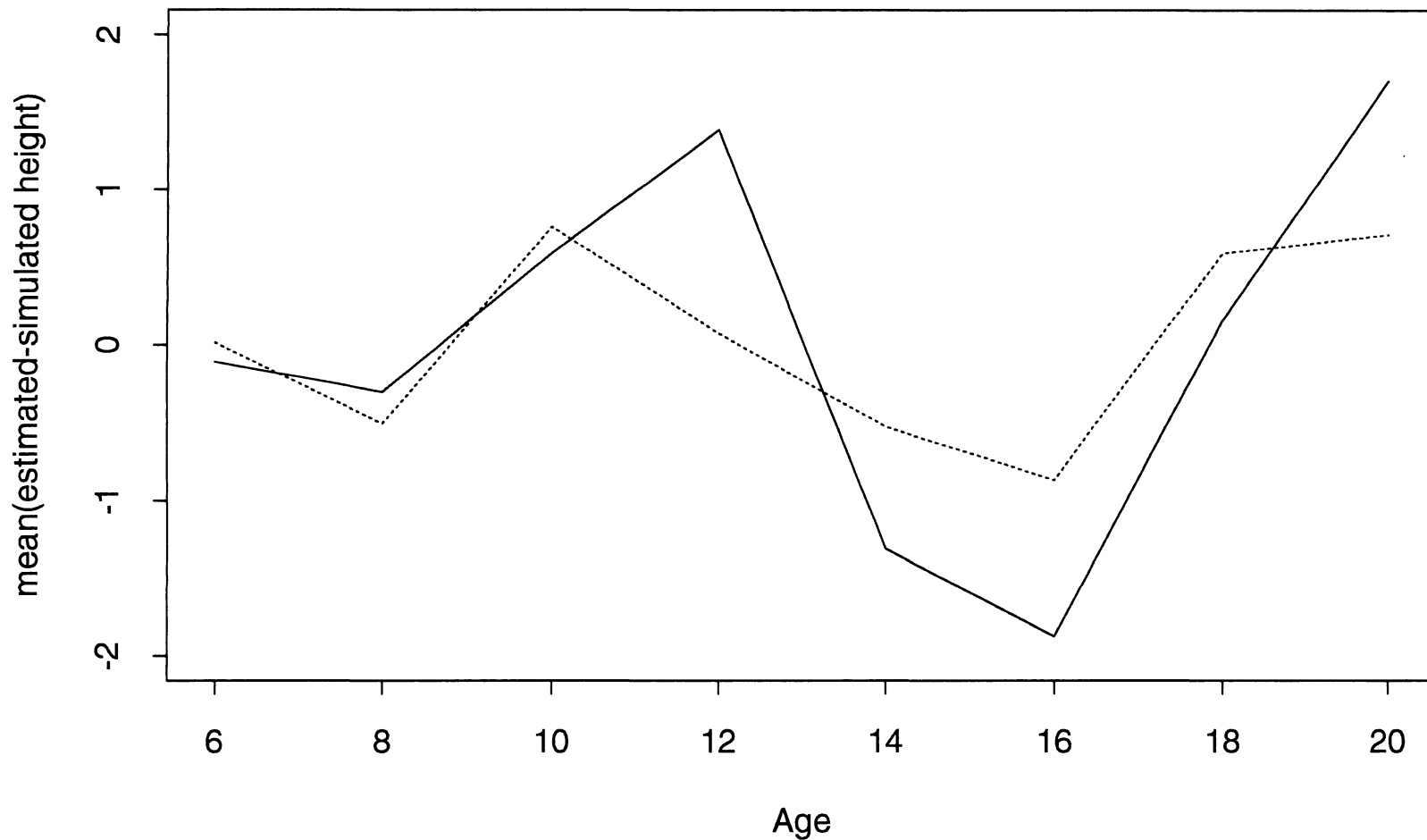


Figure 2: The mean differences between estimated and simulated human growth curves with 8 design points per curve. Notice large differences between the spline smooth and simulated curves in the teen years (solid), compared to the somewhat flatter behavior of the EB estimator (dashed) in the same region.

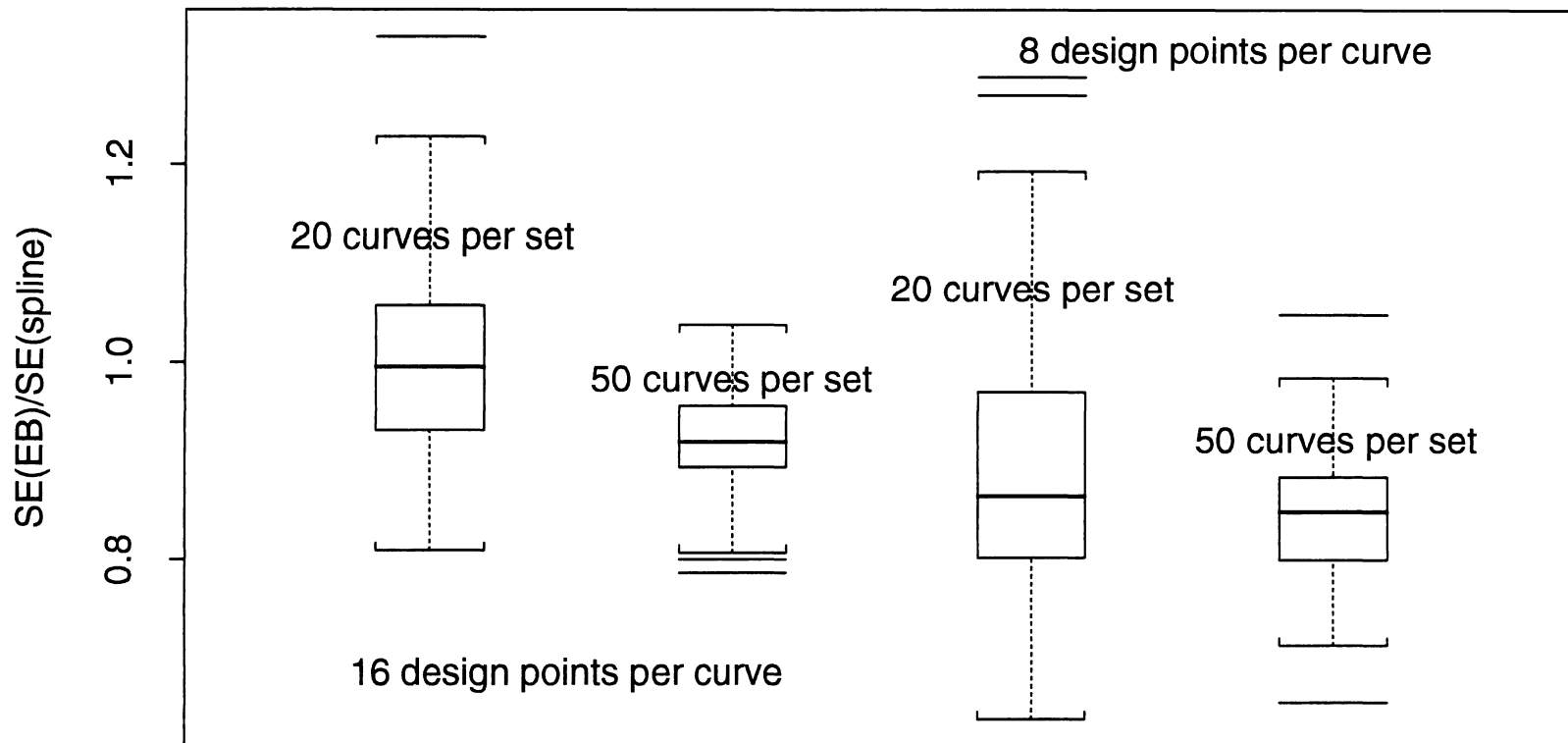


Figure 3: Each boxplot above represents the ratio of the SE of the EB estimator to the SE of the smoothing spline estimator for 100 replicates of simulated human growth curves. The more curves available for estimating the underlying population curve, the greater the improvement of the EB estimators. As well, the fewer design points per curve (and thus the less precise the spline estimator) the greater the improvement of the EB estimator.

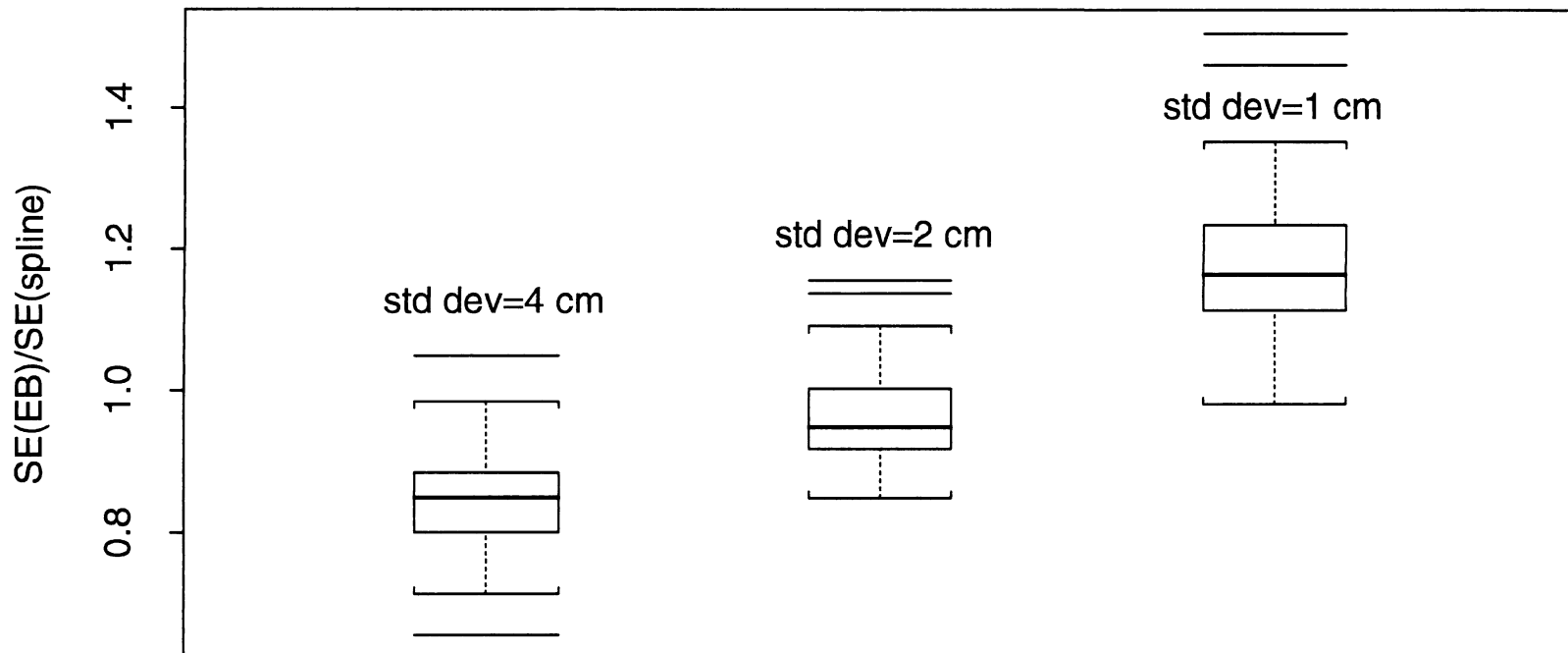


Figure 4: Each boxplot above represents the ratio of the SE of the EB estimator to the SE of the smoothing spline estimator for 100 replicates of simulated human growth curves with 8 points per curve and 50 curves per sample. The greater the standard deviation of the measurement error, the greater the improvement of the EB estimators. For this set of mean curves, several hundred curves are required to obtain improvement when the standard deviation is 1 centimeter.

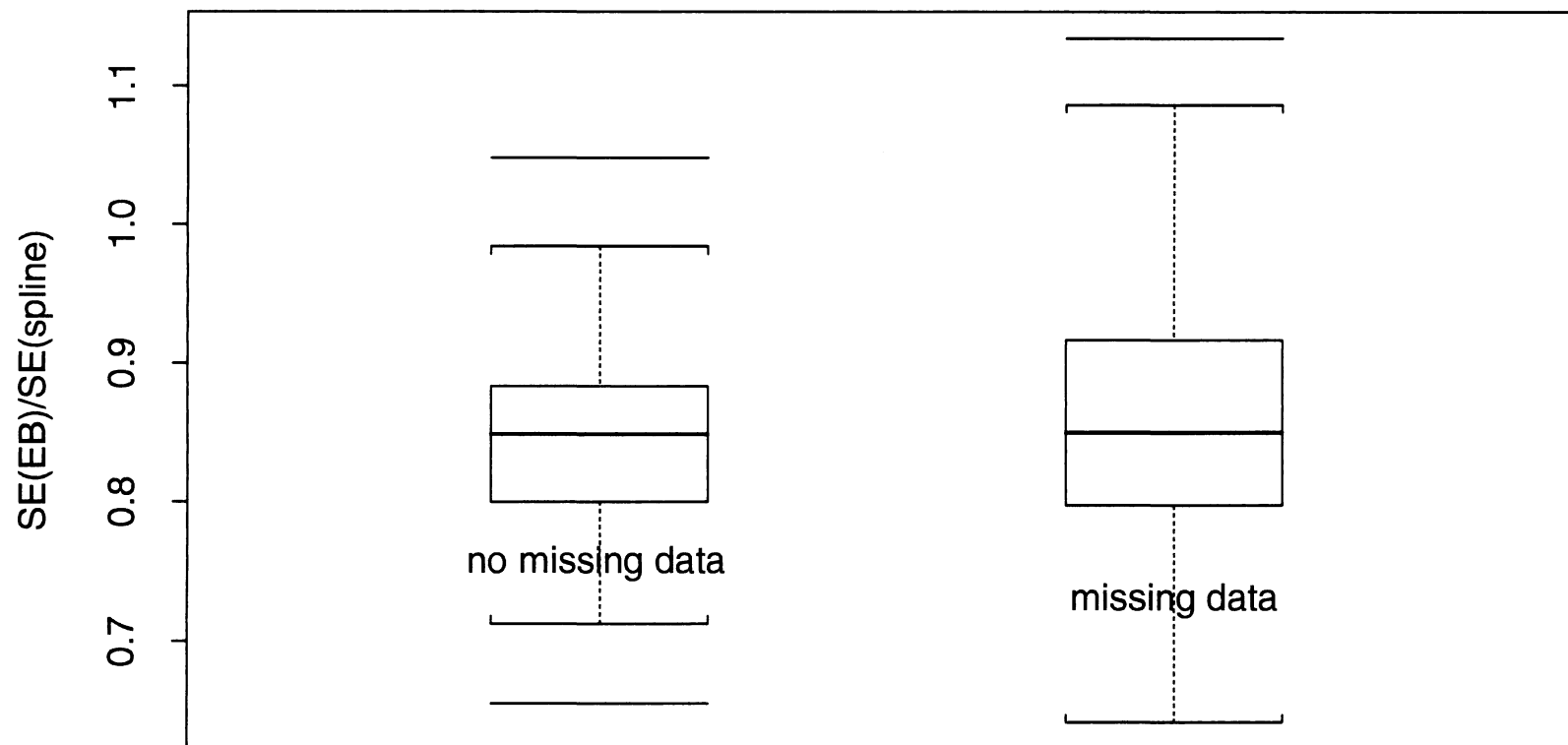


Figure 5: Each boxplot above represents the ratio of the SE of the EB estimator to the SE of the smoothing spline estimator for 100 replicates of simulated human growth curves with 8 points per curve and 50 curves per sample. The effect of missing data is to increase the spread of the ratio of SE for the EB and smoothing spline estimators, most probably due to less precise approximation of the shrinkage parameter.